機械式音声合成装置とソフトマテリアル

大 須 賀 公 一*

Mechanical Speech Synthesizer and Soft Material

Koichi OSUKA (Kobe University, 1-1 Rokkoudai-cho, Nada, Kobe 657-8501, JAPAN)

Mechanical Speech Synthesizer is a real mechanical model of human's vocal production mechanism. Our vocal mechanism is composed of "Tongue", "Lips of mouth", "Vocal chords" and "Voice road" roughly separately. To realize these components, we should use some soft materials because they are basically constructed with the soft element. In this note, firstly, we introduce a brief history of speech synthesizer, and then show some importance of development of mechanical speech synthesizer. Next, we show some developed mechanical speech synthesizers. Finally, we discuss the next generation mechanical speech synthesizer.

Key Words : Mechanical Speech Synthesizer, Artificial Tongue, Silicone Rubber, Artificial Vocal Chords

1. は じ め に

機械式音声合成装置 (Mechanical Speech Synthesizer: MSS) とは、人間の発声機構を人工的に再現した実態モデルである。人の発声機構は大きくわけて、「舌」「口唇」「声帯」「声道」から構成されている。これらは基本的には軟体要素で構築されているために、人工的に再現するためには何らかのソフトマテリアルを用いる必要がある。本稿では、機械式音声合成装置を含む音声合成に関する概観、機械式音声合成装置を開発する意義、具体的な開発例、今後の展望などについて述べてみたい。

2. 音声合成と機械式音声合成装置

我々が社会生活を営むために非常に重要な役割を果たしている「音声」は人類独特の情報媒体である。もちろん多くの動物や鳥類も何らかの発声は行うが、人間ほど複雑な発声メカニズムを持っているものはない。

そのようなことから、我々は古くから「音声の発話メカニズム」に対して大きな好奇心をもっており、なんとか人工的に再現することで理解しようとしてきた。例えば、1700年代にはケンペルンがふいごにより革製の筒に空気を送り込み、その筒の形状を手で変化させて母音と子音を合成させたとされている¹⁾。図1参照。

その後、1900年代になると、電気回路で発声を再現しようとする試みが始まる。これは、発声の物理的メカニズムを再現するのではなく、音声は、結局は空気の振動の合成であるとの考えに基づき、音として音声を再現しようとするものである。すなわち、パルスの発声回路や声道の共鳴特性を模擬する音響的なフィルターを数種類用意し、それらを電気的に駆動することで様々な振動波形が再現できるようにしておく。そうすると、その回路網のパラメータを適切に調整することによって、その信号をスピーカーに人力することで発音すれば音声に近い音が構成できる²⁾。

さらに、計算機の発達に従って、上のような電気回路で実現していたことをソフトウェアで実現することが容易に

図1 ケンペルンの機械式音声合成装置(ドイツ博物館にて撮影)

*神戸大学工学部機械工学科(〒657-8501 神戸市灘区六甲台町1-1)工学博士。1984年、大阪大学大学院基礎工学研究科制御工学専攻修士課程修了。同年、(株)東芝入社、総合研究所勤務。1986年、大阪府立大学工学部機械工学科助手、同講師、同助教授を経て、1998年、京都大学大学院情報学研究科システム科学専攻助教授、2003年、神戸大学工学研究科教授となり、現在に至る。ロボティクス、レスキュー工学、制御工学の研究に従事。

なってきたため、最近では計算機上に音声合成のためのアルゴリズムを構築することが多くなっている。

以上のような音声合成に関する簡単な歴史を踏まえ、あえてここでは最も古い流れである機械式音声合成装置について述べるわけであるが、実はその開発には現代でも大きな意義を有していることを次章ではみてゆこう。

3. 機械式音声合成装置を開発することの意義

本章では、MSSを開発することの意義を音声学、工学、臨床医学、ロボット工学・機械工学などの側面から考察する。

3.1 音声学的意義

人間の知能の解明は非常に大きなテーマである。その一つの足がかりとして発声のメカニズムの解明がある。なぜなら、発声による会話は人間にしか見られない能力であり、人間の脳における音声に関係する部分が広いからである。

発声メカニズムを解明・理解するための方法論には2つある。一つは「解析による解析(生物学的検討)」であり他は「構成による解析(工学的検討)」であり両者が適切に融合する必要がある。ここでは後者について考察を行うが、そのためには何等かの発声モデルを構築する必要がある。その有力な方法の一つが計算機内に発声機構全体のシミュレータを構築することである。ところが、これにはいくつかの問題があり、そのためにMSSを構築する必要がある。以下で説明しよう。

3.1.1 数値解析の限界

原理的には、人間の声は声帯で発生された声帯音源波が口内(声道)を伝播することで発声される。物理現象としては空気振動であり、理論的にはナビエ・ストークスの方程式を解けばよい。ただし、口内は非常に複雑な形をしており、単純な数値解法では解けない。そこで有限要素法が有効になるが、境界面の複雑な形態から、たとえば、声帯付近では1cmあたりメッシュを24000にも切る必要がある(そうでないと数値的に発散してしまう)。これは現在のスーパーコンピュータ(ギガフロップスマシン)でも1晩かかる程度である。声道は16cm程度あるので全体の詳細なシミュレーションは非常に時間がかかる。

3.1.2 声帯・舌シミュレーションの困難性からの意義

時間がかかるだけであれば、計算機が速度が向上すればいつかは解決される可能性がある。実は、それよりも本質的に困難な要因がある。声帯や舌は発声時には自由運動、衝突、めり込みを起こす。このような現象をシミュレーションしようとした場合、境界条件の決定が非常に困難でほとんど不可能であるといわれている。また、本分野と関連が深い流体工学では風洞実験はなくなると言われている。それは、シミュレーションモデルの種々のパラメータを決定するためにはどうしても実態モデルが必要だからで

ある。例えば、自動車、翼関係のシミュレーションでは自動車、翼の実態モデル(あるいは実物)があってそこから実験的に種々のパラメータを決めている。すなわち、シミュレーションモデルを構築したり、その妥当性を検証したりするために実態モデルは必要なのである。発声に関しては、自動車などの場合と異なり、生体を用いた精密な検証実験が困難なため、他の分野に比較して、実態モデル(MSS)の必要性はより大きい。また、シミュレーションモデルで表現できない複雑な現象(流体的現象や生体的現象)を再現しそれを観測するためにもMSSは必要となる。

3.2 工学的意義

実用的な音声合成装置(計算機や電子回路)ーリアルタイム性を有している小規模であるが性能の高いものーを構築するためには適切な「近似」が必要である。すなわち、実用的音声合成装置を設計するための「設計用モデル」が必要となる。これを得るためには精密なモデルである「詳細モデル」が必要である。基本的には計算機の中に構築すれば良いが、一般に人工物の場合、詳細モデルは同定実験などを行い、実物からの種々のデータ(豊富に取れる)を数理的処理を行うことによって求めることが多い。しかしながら、人間の発声器官の場合には、生体からのデータを獲得することが困難(口の中や鼻空や声帯の前後などに多くのセンサーを配置する必要があるが、それはほとんど不可能)である。そこで「詳細モデル」構築のために詳細な実態モデルが必要となる。これがMSSである。

3.3 臨床医学的意義

口腔外科などで、下顎等を損傷した患者に発声方法(舌や唇の動かし方や筋肉の運動)を教示する際に実態モデルが欲しいという要求がある。絵やCGよりも3次元上の実物を用いて説明する方がよいという考えである。リハビリでも同様である。

3.4 ロボット工学・機械工学的意義

口唇・舌などは軟体機械であり(分布アクチュエーター系・超多自由度系)その制御問題は興味深い。ヒューマノイドの構築を考えると発声機構も人間と同様のものを開発することが自然である。

3.5 考察

以上、本章では、いろいろな面から見た機械式音声合成装置(MSS)の開発の意義を考察した。実際に開発する過

図2 機械式音声合成装置(形態模写モデル)

程で考慮しなくてはならないのは「形態の模倣」を目指すのか「機能の模倣」を目指すのか(あるいは両方なのか)ということである。音声学や臨床医学的には「形態の模倣」が重要になる。一方、工学的には「機能の模倣」で十分である。例えば、人類が鳥をみて同様の機械を作ろうと思ひ、「形態の模倣」である羽ばたき機ではなく「機能の模倣」である飛行機を開発したと同様である。

4. 機械式音声合成装置の実現にむけて

本章では、筆者らがこれまで開発してきた機械式音声合成装置について紹介してゆく。

4.1 形態模倣

まず、発声器官(声帯音源波、声道、舌、口唇など)の物理的な形状の模倣ができれば、音声の発声が可能であることをたしかめるために、図2のような模型を試作した³⁾。左が全体写真であり、口唇を含めた頬のマスクが見える。また下の方には声帯音源波(この場合には三角波を用いた)を発音するためにスピーカーがある。

このモデルは、人間の発声に関わる種々の機関をできるだけ忠実に再現しようとしてもので、基本的には石膏とシリコーンゴムで実現されている。具体的には図3左に示すように、上下顎の再現(石膏製)と図3右に示すような声道(シリコーンゴム製)の再現も行っている。

この実験の目的は、声道や口唇の形状、舌形状が適切に構成されていれば、人間と同様に例えば母音は発声できるということを確認することであった。そのために、実験者が日本語5母音を発声する時の頬と口唇の形を模倣した石膏のマスクを試作し(その形状模倣は鏡を用いて実験者自らが行った)、さらに、舌形状も想定する母音が発声できるように試行錯誤で形状を定めていった。その結果を図4と図5に示す。

舌部形状の具体的な決定は次のような手順で行った。まず、合成したい母音のマスクを装着させ、声帯音源波に相当する三角波信号を下のスピーカーに注入する。そして、合成音声のワースペクトル形状(合成波形)が目標となる実際に人間が発声した音声のパワ

ースペクトル波形(目標波形)に近づくように硬化前の紙粘土からなる舌部の形状を手で変形させる。その際、できるだけ試行錯誤をへらすために、舌形状の初期形状は、文献4に示されている正中断面での舌形状を参考に与えた。合

図3 上下顎部(左)と声道部(右)

図4 推定された口唇形状

図5 推定された舌形状

成波形が目標波形に近づいた時点で舌部を硬化させる。

図6にその実験結果の一例を示している。各図の横軸は周波数で縦軸はパワースペクトルの目標形状(肉声音声：細い線)と合形状(合成音声：太い線)である。一般に、人間が発声した母音はフォルマント周波数(声道の共鳴周波数)のうち周波数の低い第1, 第2フォルマントによってほぼ特徴付けられると言われている。もちろんその2つの情報のみならず、パワースペクトルの全体的な形状も重要な役割を果たしている。

以上のような観点から図6に示した実験を見るとほぼ、合形状は目標形状に近づいていることがわかる。実際、発声音は、ほぼ日本語5母音に聞こえた。

本章の結果から、ある意味で当然ではあるが、確かに種々の音声発声器官の形状を模倣すれば、人間と同様な発声が可能であることがわかった。そこで次のステップとして、

各要素の詳細な再現を考えた。すなわち、先の実験では声帯はスピーカーからの信号であったり、口唇、舌などはそれぞれの母音に応じた固定された要素を用いていたのに対し、それらを実際の声帯と同様の実態モデルからの音にする、口唇や舌を駆動可能な実態モデルに置き換えるという試みを行った。

4.2 声帯と口唇

人間の声帯は、発生学的には、肺に外から異物が入らないように気管につけられた蓋である。従って、軟体構造物で構成されており、筋肉によって開閉ができる。人はその構造を利用して音声を発声するための音源にしている。それを人工的に再現するにはシリコンゴムが適している。図7に筆者らが開発したシリコンゴムで作られた人工声帯を示す。構造は非常に簡単で直径約13mmのプラスチック管の先端に図のようなシリコンの膜を固定する。そして管の他方から気流を流し込む。そうすると、シリコンゴムの隙間に乱流が発生し、自励振動が発生する。それが声の元である声帯音源波となる。

また、口唇を発声によって変形させる必要がある。

図7 人工声帯

図8 口唇モデル

図6 発声実験結果

図8は筆者らが試作した口唇モデルである。これはシリコンゴムで作られており、皮膚と骨格の間にワイヤー駆動の筋が内蔵されている。これらを駆動させることで日本語5母音を発声する際の口唇の形状が再現できるようになる。

4.3 舌構成

機械式音声合成装置の開発で最も困難なのは舌の再現である。舌は、骨のような固い構造体を有しておらずすべて柔らかい筋肉で構成されており、その点がいわゆる従来からの機械とは全く異なっている。言い換えれば、舌は、アクチュエーターと構造体が一体化した柔らかい要素から構成されていると言える。

そこで筆者らは、柔らかいアクチュエーターとして、図9のようなシリコンゴムで作られた蛇腹状のチューブWorm Tubeを開発し、それを用いて舌の形状再現を試みた⁵⁾。

Worm Tubeは、図9のように伸縮可能な可縮部(外径約9mm)と接続チューブから成り立っており、接続チューブから空気を抜くと図上右のように伸縮する。その動作を筋肉に見立てて舌を構成した。すなわち、種々の長さのWorm Tubeを用意し、図10左のような人の舌の筋肉配置を参考にして人工舌におけるWorm Tubeの配置を定めた(図10右)。

そして、各Worm Tubeは市販の小型注射器とシリコンチューブで接続されており、その注射器はステッピングモーターで伸縮するようになっている。そのステッピングモーターを計算機で制御することでWorm Tubeの長さを制御する。このようなセットを13セット用意し、舌全体を構成した。本システムをBERO(Biomimetic Tongue ROBOT)と名付けた。

図12にBEROを用いた舌形状の再現の様子を示す。先に述べた13本のWorm Tubeの長さ(注射器による空気圧の調整)を適切に行うことによって、日本語5母音を発声する際の舌形状が再現できていることがわかる。

Worm Tubeの構造体としての剛性が不足していることから三次元舌形状の再現までには至っていない。今後は、Worm Tubeの形状、構造なども含めてより詳細な検討が必要である。

5. おわりに

本稿では、音声合成の中でも機械式音声合成装置に着目し、その開発意義、筆者らが開

発してきた種々の装置について概観した。

人間は、声帯からの声帯音源波が舌や上下顎、口唇などで形成される声道を通過する際の共鳴現象を利用して発声している。すなわち、発声時の声道の3次元的な形状が人工的に再現できれば同様な発声は可能になる。そのような観点から機械式音声合成装置を開発する流れもある⁶⁾。それに対して筆者らは、舌の筋肉配置なども考慮した軟体機械としての舌の構築を目指している。結果として、声道の形状制御ができるという立場である。そのようなことから舌を小さな三角要素の集合体として構成する方法(いわば有限要素法の実態版)も行っている⁷⁾。

最後に、筆者らの一連の研究は、CRESTの「発声力学

図9 Worm Tube

図10 舌の筋肉配置と人工舌におけるWorm Tubeの配置

図11 BEROの構成図

図12 BEROによる日本語5母音舌形状の再現

に基づくタスクプランニング機構の構築」というプロジェクトの一環として行われたことを付記しておく⁸⁾。

Project (1998-2002)

<http://www.brl.ntt.co.jp/cs/speech/crest-j.html>

References

- 1) Honda, K.: Tsushin Sogokenkyusho Kiho, **47-3**, 43 (2001)
- 2) Okada, M.: "Kutigomoru Computer, Jyoho Frontier Series", Jyoho Syori Gakkai Ed., Kyoritsu, Tokyo, Chapter 1 (1995)
- 3) Osuka, K., Araki, Y., Sawada, K., Ono, T.: *Nippon Robot Gakkaishi*, **16-2**, 189 (1998)
- 4) Chiba, T., Kajiyama, M.: Phonetuc Society of Japan, Chapter 3 (1958)
- 5) Sawada, K., Osuka, K., Ono, T.: *Nippon Robot Gakkaishi*, **17-7**, 1001 (1999)
- 6) Nishikawa, T., Kuwae, T., Akiyama, T., Fukui, K., Takanobu, H., Mochida, T., Homareda, M., Kousai, A.: *Denshi Jyoho Tsushin Gakkai Kenkyuhokoku*, **103-749**, 37 (2004)
- 7) Bando, M., Osuka, K.: *Robotics Mechatronics Koenkai '04 Yoshisyu*, 1A1-L1-29 (2004)
- 8) Kagakugizyutsu Shinko Jigyodan Kisokenkyu Suishinjigyo (CREST Proj.) Kenkyuryoiki "Nou wo tsukuru" Kenkyu

日本語表記参考文献

- 1) 本多清志：通信総合研究所季報，**47-3**，43 (2001)
- 2) 岡田美智男：口ごもるコンピュータ，情報フロンティアシリーズ，情報処理学会編，共立出版，東京，1章 (1995)
- 3) 大須賀公一，荒木祐之，澤田謙次，小野敏郎：日本ロボット学会誌，**16-2**，189 (1998)
- 5) 澤田謙次，大須賀公一，小野敏郎：日本ロボット学会誌，**17-7**，1001 (1999)
- 6) 西川員史，桑江俊治，秋山隆之，福井孝太郎，高信英明，持田岳美，譽田雅彰，高西淳夫：電子情報通信学会研究報告，**103-749**，37 (2004)
- 7) 板東幹雄，大須賀公一：ロボティクス・メカトロニクス講演会 '04，1A1-L1-29 (2004)
- 8) 科学技術振興事業団 基礎研究推進事業 (CREST Project) 研究領域「脳を創る」(統括 甘利俊一) 研究プロジェクト (1998-2002)，発声力学に基づくタスクプランニング機構の構築 <http://www.brl.ntt.co.jp/cs/speech/crest-j.html>

*

*

*

*

*